Discussion Paper Series

2004 – 19

Department of Economics
Royal Holloway College
University of London
Egham TW20 0EX

Is more data better?*

Kaushik Mitra
Royal Holloway College, University of London

June 28, 2003; revised

Abstract

Conventional wisdom usually suggests that agents should use all the data they have to make the best possible prediction. In this paper, it is shown that agents may make better predictions by discarding old data if their model is mis-specified. The applicability of the results to some economic models is also demonstrated.

Journal of Economic Literature Classification Numbers: C13, C22, C53, D83, E32, E37.

Keywords: optimal, mean squared error, bounded memory.

Corresponding address: Department of Economics, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, U.K. Email: Kaushik.Mitra@rhul.ac.uk, Phone: +44 1784 44 3910, Fax: +44 1784-43 9534.

1 Introduction

The starting point of the paper is the presumption that if the true data generating process is unknown (and potentially complex), economic agents can

***Acknowledgements**: This paper was written when I was a postdoctoral fellow at the Research Unit on Economic Structures and Growth (RUESG), University of Helsinki. Financial support from the Academy of Finland and the Yrjö Jahnsson Foundation is gratefully acknowledged. I am very grateful to Seppo Honkapohja for asking me the question which led to this paper. I thank George Evans for helpful discussions and Karim Abadir, Emilio Barucci, John Duffy, Ed Greenberg, Cars Hommes, Manuel Santos, Gerhard Sorger and Gabriel Talmain for their comments. Comments from two referees have substantially improved the exposition of the paper. The usual disclaimer applies.

be expected to use simple underparametrized representations of the process to make their forecasts. They can then obtain the best forecast within this class. In the terminology of Sargent (1999, Ch. 6), agents have "optimal mis-specified beliefs". This paper considers a similar situation of mis-specified forecasting.

The broad idea is as follows. Suppose, the variable of interest to the economic agent (like price of a good) follows an autoregressive process of order k, i.e., an AR(k) process. It is possible that agents may get the form of this process right (say, through some specification tests) and estimate the parameters of this process which they then use to forecast the next period value of the variable. Standard econometric theory then tells us that with large enough data, the estimated parameter values will converge to the true values. However, it is not clear what will happen if the true process is unknown or if the agent fails to arrive at the correct form of the process. In these situations, it is conceivable that agents may use a simple ("rule of thumb") method for making the short-term forecast. This simple rule would be plausible in some cases. Another scenario may be that the agent under-parametrizes the AR(k) process to be an AR(l) process; $k > l$. This situation may seem particularly realistic if k is large. In Sargent's spirit, the agent can then choose the parameters of the AR(l) process optimally (using some objective criterion) and form the forecast.

In order to get new analytical results, I focus on the simplest situation where the true process is AR(1). The form of this process is not known to the agent and he deals with this lack of knowledge by using the sample mean of the previous T data points (T is referred to as the memory length) to make the forecast. It seems that computing a sample average is a reasonable first approach to making the forecast. This rule also arises from the simplest form of mis-specification in this set-up: agents mis-specify (under-parametrize) the order of the process to be AR(0), that is, a constant plus some white noise, and hence use the sample mean of data to form the forecast. However, T is chosen optimally to minimize the one period ahead forecast error. This simple mis-specification enables us to get (to the best of my knowlege) some new results for this forecasting problem. The intuition gained from here will hopefully continue to be relevant for more complicated scenarios.

We have so far presented the basic idea in an abstract set-up. An equally important aim of the paper is to present some concrete economic example models where the scenario considered becomes directly relevant. Section 3 describes some of these models.

2 The Main Result

Assume that a random variable μ_t evolves according to a first order autoregressive process (AR(1)) as specified below

$$\mu_{t+1} = \lambda\mu_t + (1-\lambda)\bar{\mu} + \varepsilon_t; 0 \leq \lambda < 1. \tag{1}$$

where $\{\varepsilon_t\}$ is white noise with $E\varepsilon_t = 0, E\varepsilon_t^2 = \sigma_\varepsilon^2$ and μ_0 is given. The unconditional (asymptotic) mean of the μ_t process is given by the constant $\bar{\mu}$. The true data generating process for μ_t is assumed unknown to the agents but they need to forecast the current value of μ_t. At time t, this forecast, $\mu_t^e(T)$, is given by

$$\mu_t^e(T) = T^{-1}\sum_{i=1}^{T}\mu_{t-i} \tag{2}$$

i.e., by the arithmetic mean of the previous T data points, $\{\mu_{t-1}, \mu_{t-2,} .., \mu_{t-T}\}$. T is called the memory length of the agent. The forecast error made at t is $\mu_t^e(T) - \mu_t$. We assume that the process has been running for a long period of time so that $t \to \infty$ gives a reasonable approximation of this process. This approximation makes the moments of the forecast error independent of the initial condition of the process. This is in line with what is done in econometrics: one is interested in the statistical properties of estimators or predictors in the long run, i.e., once the influence of the initial conditions has died down.

Denote the mean squared error (MSE) of $\mu_t^e(T), E[(\mu_t^e(T) - \mu_t)^2]$, by $MSE_t^{est}(T)$ and the *asymptotic* MSE, $\text{limit}_{t\to\infty}\, MSE_t^{est}(T)$ by $MSE_\infty^{est}(T)$. The basic problem is to determine the memory length T of the forecast (2) which minimizes $MSE_\infty^{est}(T)$. In other words, the problem is to determine the optimal memory length T for predicting the next realization of an AR(1) process by the arithmetic (sample) mean of the last T observations.

Under rational expectations (RE), agents are assumed to know the true data generating process, an assumption which is usually considered implausible. In this model agents deal with their lack of knowledge of the true structure by using the simple rule (2). However, this rule has much to be said in its favor. For one thing, it is asymptotically unbiased for all memory lengths T, i.e., $\text{limit}_{t\to\infty} E(\mu_t^e(T) - \mu_t) = 0$, for all T. Secondly, the law of large numbers of Markov processes implies that with large enough data the forecast is expected to converge to the true mean of the asymptotic distribution of the process. Thirdly, the forecast $\mu_t^e(T)$ encompasses the optimal prediction for the important borderline cases of an i.i.d sequence (when $\lambda = 0$) and a random walk world (when $\lambda = 1$). If $\lambda = 0$, $T \to \infty$ is optimal for prediction whereas $T = 1$ is optimal when $\lambda = 1$.

We now explore whether the optimal T is affected when λ is between 0 and 1. As a preliminary step we prove the following proposition.

Proposition 1 *For any* $\lambda \in [0, 1)$, *we have*

$$MSE_{\infty}^{est}(T) = \sigma_{\varepsilon}^2 [\frac{(1-\lambda)^2 T(T+1) \ + \ 2(1-\lambda)\lambda^{T+1}T \ - \ 2\lambda(1-\lambda^T)}{(1-\lambda)^3(1+\lambda)T^2}] \quad (3)$$

Proof. See Appendix A. ■

When $\lambda = 0$, $MSE_{\infty}^{est}(T)$ clearly decreases monotonically with T. However, it is not obvious what happens when $\lambda > 0$. To understand this, I prove the following proposition.

Proposition 2 *For any* $\lambda \in (0, 1)$, $MSE_{\infty}^{est}(T)$ *decreases monotonically with* T *for all* $T \geq T(\lambda) \equiv 4\lambda(1-\lambda)^{-2}$.

Proof. See Appendix B. ■

Clearly $T(\lambda)$ increases monotonically with λ. For example, $T(.25) \approx 2$ and $T(.9) = 360$. Proposition 2 shows that $\mathrm{MSE}_{\infty}^{est}(T)$ decreases monotonically with T for all $T \geq 1$ provided λ is *small* enough so that $T \to \infty$ is optimal. In fact, we can prove that this is true for all $\lambda \in [0, 0.5]$ as shown in the following proposition.

Proposition 3 *For any* $\lambda \in [0, 0.5]$, $T \to \infty$ *minimizes* $MSE_{\infty}^{est}(T)$.

Proof. See Appendix C. ■

This leaves unanswered what happens when $\lambda > .5$? We are at least able to provide a partial answer to this question in the following proposition.

Proposition 4 *For any* $\lambda \in (.5, .88]$, $T = 1$ *minimizes* $MSE_{\infty}^{est}(T)$.

Proof. See Appendix D. ■

The proof of Proposition 4 may lead one to suspect that $T = 1$ is optimal for all $\lambda \in (.5, 1)$. One can, in principle, look at values of λ arbitrarily close to 1 and solve the corresponding polynomial inequalities. However, the computation time increases rapidly. Instead I computed the MSE numerically for values of λ close to 1 and this leads me to the following conjecture.

Conjecture 5 *The optimal T is 1 for all $\lambda \in (.5, 1)$.*

One can provide some intuition for these results. λ may be interpreted as the degree of mis-specification given the agents' beliefs about the data generating mechanism. If agents believe they live in an i.i.d world ($\lambda = 0$), it is optimal to use $T \to \infty$. If the true λ is close to 0, $T \to \infty$ continues to be optimal since the model mis-specification is not very severe. However, if the true λ is close to 1, then the mis-specification is very severe and it is no longer optimal to use $T \to \infty$. Similarly, if agents believe in a random walk world, it is optimal to use $T = 1$ (the best prediction is given by the last realization). This continues to be optimal when λ is close to 1 since the model is then not too mis-specified but is no longer optimal when the model is heavily mis-specified, say, for λ close to 0.

The next section describes some economic example models where the results of this section can be applied.

3 Economic example models[1]

3.1 Profit maximization by the firm

Example 1. This example follows Muth (1961). Consider the problem of a firm choosing output in every period t based on its forecast of the market price. The realized price in period t, p_t, follows the exogenous stochastic process (1) ($p_t \equiv \mu_t$).[2] This assumption would be appropriate in an open economy or for a monopolist facing infinitely elastic demand or for a firm producing in a competitive market. The firm chooses output q_t at the end of period $t-1$ to maximize expected period t profits. Assuming quadratic costs $cq_t^2/2$, profits (Π_t) are given by

$$\Pi_t = p_t q_t - cq_t^2/2$$

so that expected profits are maximized by choosing $q_t = c^{-1}p_t^e$ where p_t^e is the forecast of p_t of the firm at the end of period $t-1$. In our case, p_t^e (or $p_t^e(T)$) is given by (2) with $p_t = \mu_t$. By using the optimal choice $q_t = c^{-1}p_t^e(T)$, profits

[1] The first two examples are borrowed from Evans and Ramey (1998) and the third one is from Honkapohja and Mitra (2003).

[2] However, ε_t has a bounded support to ensure the non-negativity of price.

may be rewritten as

$$\Pi_t(T) = (2c)^{-1}(2p_t p_t^e(T) - p_t^e(T)^2)$$

and the firm chooses T to maximize expected profits, $E\Pi_t(T)$.

Suppose, on the contrary, the firm chooses T to minimize the MSE of $p_t^e(T)$, $E[p_t^e(T)-p_t]^2$. Since $E[p_t^e(T)-p_t]^2 = E[p_t^e(T)^2 - 2p_t p_t^e(T) + p_t^2]$, and p_t is exogenous, this is equivalent to choosing T to minimize $E[p_t^e(T)^2 - 2p_t p_t^e(T)]$ i.e., to maximize $E\Pi_t(T)$. We assume that the price process has been running for a long period of time so that it is appropriate for the firm to maximize $E\Pi_t(T)$ as $t \to \infty$.

3.2 Permanent Income Hypothesis

Example 2. This corresponds to the first example in Lucas (1976). Consumption is given by

$$\begin{aligned} c_t &= c_{pt} + u_t, \\ c_{pt} &= k y_{pt}, \\ y_{pt} &= (1-\delta)\sum_{i=0}^{\infty} \delta^i y_{t+i}^e, \ \ 0 < \delta < 1. \end{aligned}$$

Here u_t is a white noise process denoting transitory consumption. c_{pt} denotes permanent consumption, y_{pt} is permanent income, δ is the household's discount factor and y_{t+i}^e is the household's time t forecast of income at time $t+i$, y_{t+i}. The income process y_t is assumed to follow an AR(1) process ($y_t = \mu_t$ in (1)). While Lucas (1976) assumed rational expectations, we instead assume that $y_{t+i}^e = y_t^e(T)$ for all $i = 0, .., \infty$ so that $y_{pt} = y_t^e(T)$ and $c_t = k y_t^e(T) + u_t$. As before, $y_t^e(T) = T^{-1}\sum_{i=1}^{T} y_{t-i}$. The agent's problem is to choose the memory T which minimizes $\text{limit}_{t\to\infty} E[y_t^e(T) - y_t]^2$ so that the results of Section 2 apply. Note that agents use the same forecast, $y_t^e(T)$, for making predictions over longer horizons. This may be rationalized either by assuming that agents believe (approximately) in an i.i.d world (Abraham and Ledolter (1983)) or in a random walk world (Hamilton (1994)).

3.3 The Muth Market Model

Example 3. This extends Example 1. Assume a competitive market model (the Muth market model) with identical suppliers. The optimal supply of a representative supplier from Example 1 is given by $q_t = c^{-1}E^*_{t-1}p_t$ (we use the notation $E^*_{t-1}p_t$ instead of p^e_t now). $E^*_{t-1}p_t$ denotes the (possibly) subjective expectations of agents whereas the same notation without the superscript *asterisk* denotes rational expectations. Suppose that actual total output is equal to the sum of optimal outputs plus an iid shock v_t with zero mean. This yields the market supply function

$$q^s_t = c^{-1}E^*_{t-1}p_t + v_t,$$

to which we append a (downward sloping) demand function

$$q^d_t = C - Bp_t.$$

Here, $q^j_t, j = d, s$, denote aggregate quantities demanded and supplied and B, $C > 0$. Using equality of supply and demand, we obtain

$$p_t = \alpha + \lambda E^*_{t-1}p_t + u_t, \tag{4}$$

where $u_t = -B^{-1}v_t$, $\alpha = B^{-1}C$ and $\lambda = -B^{-1}c^{-1}$. This model has a unique rational expectations equilibrium (REE), $p_t = \bar{a} + u_t$, where $\bar{a} = \alpha/(1-\lambda)$. Now assume that agents do not have RE and instead think that they are in this steady state but do not to know the value of the constant $\bar{a}$. In accordence with the macroeconomic learning approach, agents have a perceived law of motion (PLM) of the form $p_t = A + u_t$, and a natural estimate for A is then given by $E^*_{t-1}p_t = T^{-1}\sum_{i=1}^{T} p_{t-i}$.[3] Using this in (4) yields an $AR(T)$ process for p_t when all agents use the same T, i.e.,

$$p_t = \alpha + \frac{\lambda}{T}\sum_{i=1}^{T} p_{t-i} + u_t. \tag{5}$$

Assume that all agents use $T = 1$ in their forecasting so that p_t is an $AR(1)$ process. Does any single agent have an incentive to use $T > 1$, when all other agents use $T = 1$? To answer this, a single agent can find the value of T that minimizes $\lim_{t\to\infty} E[T^{-1}\sum_{i=1}^{T} p_{t-i} - p_t]^2$. From Section 2, $T = 1$ minimizes this expression if $1 > \lambda > 0.5$. In this sense, $T = 1$ is a *self-confirming equilibrium* in memory lengths if $1 > \lambda > 0.5$.[4]

[3] See Evans and Honkapohja (2001) for an extensive discussion of the macroeconomic learning approach.

[4] This terminology has been used in Sargent (1999). The term *restricted perceptions equilibrium* is also sometimes used, see Evans and Honkapohja (2001, ch. 13-14). We also note that in the basic Muth model $\lambda < 0$ but there are extensions of the model leading to $\lambda > 0$ in the reduced form; see Honkapohja and Mitra (2003) for the details.

3.4 Models with Lags

Example 4. The final example is borrowed from Evans and Honkapohja (2001, Section 13.1.2 of ch. 13). We consider the class of models given by

$$y_t = \alpha + \lambda E_t^* y_{t+1} + \delta y_{t-1} + u_t, \tag{6}$$

In (6), the market outcome depends on the agents' forecast of future values as well as lags in the endogenous variable− the reader is referred to Evans and Honkapohja (2001, Section 8.6.2 of ch. 8) for various models covered by (6). Under RE, there are typically two minimum state variable (MSV) solutions of the form

$$y_t = a + by_{t-1} + du_t$$

Assume that, as in Evans and Honkapohja, agents estimate a simple underparametrized model $y_t = a + \varepsilon_t$, i.e., they believe that the economy is in a stochastic steady state (ε_t is white noise). Hence, as before, agents forecast $E_t^* y_{t+1}$ by $T^{-1} \sum_{i=1}^{T} y_{t-i}$. Using this, the actual model is as an AR(T) process, i.e.,

$$y_t = \alpha + \lambda T^{-1} \sum_{i=1}^{T} y_{t-i} + \delta y_{t-1} + u_t.$$

In particular, if all agents use $T = 1$, then y_t becomes an AR(1) process, namely,

$$y_t = \alpha + (\lambda + \delta) y_{t-1} + u_t$$

The analysis in Example 3 is applicable: $T = 1$ is a self-confirming equilibrium provided $0.5 < \lambda + \delta < 1$.

4 Discussion and Concluding Remarks

As mentioned in the Introduction, if the true data generating process is unknown, economic agents may be expected to use simple underparametrized representations of the process to make their forecasts. A similar idea has been explored in this paper and illustrated in the examples of Section 3. This idea is related to some recent studies in the macroeconomic learning literature; these include Sargent (1999) and Evans and Honkapohja (2001). In the learning literature the specification of the agents' learning rule comes from the underlying rational expectations equilibrium (REE) of the economy. For instance, if the macroeconomic model has a REE which takes the form of an i.i.d sequence, then the agents' learning rule (the perceived law of motion (PLM)) would also take the form of an i.i.d sequence with an unknown mean, as illustrated in Example 3. The forecasting rule (2) then arises quite naturally since computing the sample average of past data is the recommended statistical procedure for estimating the unknown mean. More generally, as in Example 4, there is no reason why agents should have a PLM of the same form as the REE. Such mis-specifications may also be relatively difficult to detect. Hommes and Sorger (1998) introduced the concept of consistent expectations equilibria (CEE) by the property that the PLM and ALM are indistinguishable in terms of the sample average and autocorrelations of the observed variable. An early version of the paper had examined that if producers use the optimal T in their forecasting in Example 1, for instance, expectations are consistent in most cases.

The question of optimal memory length examined in the paper is also related to the analysis in Evans and Honkapohja (1993) and Sargent (1999). It is often suggested that agents should use a "constant gain" instead of "decreasing gain" in their learning algorithm when they suspect some structural change. This means that agents should put constant weight to current data instead of decreasing weight as in least squares estimation. This procedure helps in adapting to an exogenous time-varying process. Evans and Honkapohja (1993) examine the optimal gain parameter to use for an agent in the context of an overlapping generations economy and furthermore whether there exists equilibria where no agent has an incentive to deviate from his choice of the gain (parameter) given the gain (parameter) of all other agents (see also Evans and Honkapohja (2001, ch. 14)).[5] The self-confirming equilibrium in $T = 1$ examined in Examples 3 and 4 are obviously in the same spirit.

The current paper differs from the existing literature in that it emphasizes the size of memory to be an important issue. The memory length may affect the stability properties of REE under learning dynamics as illustrated in Example 3 and considered in detail in Honkapohja and Mitra (2003). Convergence to the unique REE takes place in Example 3 if agents use infinite

[5] Sargent (1999) and Evans and Ramey (2001) examine similar ideas.

memory when $\lambda < 1$ whereas non-convergence obtains if they use bounded memory.[6]

We feel that there are several reasons why the use of a limited data set in economic models is of interest. First, as shown here, economic agents may rationally decide to use only a limited data set if all other participants in the economy also do so. In addition, there is a certain naturalness to the assumption of bounded memory since economic data is after all limited. Perhaps no less important is the observation that econometricians seem to systematically discard old data. Finally, the dynamics under bounded memory learning has some attractive properties which the RE solution does not possess. For instance, the model of Example 3 displays excess volatility compared to the RE benchmark (see Honkapohja and Mitra for the details). As is well known, some markets (like financial markets) are indeed characterized by excess volatility.

A Proof of Proposition 1

Observe that for all i, $0 \leq i \leq T-1$, we can write

$$\mu_{t-i} = \lambda^{T-i}\mu_{t-T} + \bar{\mu}(1-\lambda)\sum_{j=0}^{T-i-1}\lambda^j + \sum_{j=0}^{T-i-1}\lambda^j \varepsilon_{t-i-j-1}$$

and hence

$$\begin{aligned}\sum_{i=1}^{T}\mu_{t-i} &= \mu_{t-T}(\frac{1-\lambda^T}{1-\lambda}) + \bar{\mu}(1-\lambda)(\frac{(1-\lambda)T-1+\lambda^T}{(1-\lambda)^2}) + \varepsilon_{t-T}\{\frac{(1-\lambda^{T-1})}{(1-\lambda)}\} \\ &\quad + \varepsilon_{t-T+1}\{\frac{(1-\lambda^{T-2})}{(1-\lambda)}\} + .. + \varepsilon_{t-3}\{\frac{(1-\lambda^2)}{(1-\lambda)}\} + \varepsilon_{t-2}\{\frac{(1-\lambda)}{(1-\lambda)}\}.\end{aligned}$$

Then

$$\begin{aligned}\mu_t^e(T) - \mu_t &= (\mu_{t-T} - \bar{\mu})\{\frac{1-\lambda^T-(1-\lambda)\lambda^T T}{(1-\lambda)T}\} + \varepsilon_{t-T}\{\frac{(1-\lambda^{T-1})}{(1-\lambda)T} - \lambda^{T-1}\} \\ &\quad + \varepsilon_{t-T+1}\{\frac{(1-\lambda^{T-2})}{(1-\lambda)T} - \lambda^{T-2}\} + ... + \varepsilon_{t-2}\{\frac{(1-\lambda)}{(1-\lambda)T} - \lambda\} - \varepsilon_{t-1}.\end{aligned}$$

[6] In the Honkapohja and Mitra setting with bounded memory, learning is asymptotically unbiased for all memory lengths, see Honkapohja and Mitra, Proposition 3, for the details.

so that

$$MSE_t^{est}(T) = \{\frac{1-\lambda^T-(1-\lambda)\lambda^T T}{(1-\lambda)T}\}^2 E[(\mu_{t-T}-\bar{\mu})^2] + \sigma_\varepsilon^2[\{\frac{(1-\lambda^{T-1})}{(1-\lambda)T}-\lambda^{T-1}\}^2 + \{\frac{(1-\lambda^{T-2})}{(1-\lambda)T}-\lambda^{T-2}\}^2 + .. + \{\frac{(1-\lambda)}{(1-\lambda)T}-\lambda\}^2 + 1)].$$

Since μ_t is an AR(1) process, $limit_{t\to\infty} E[(\mu_{t-T}-\bar{\mu})^2] = \sigma_\varepsilon^2(1-\lambda^2)^{-1}$. Consequently, as $t \to \infty$, the expression for $MSE_t^{est}(T)$ simplifies to (3).

B Proof of Proposition 2

The derivative of $MSE_\infty^{est}(T)$ with respect to T is

$$A(1-\lambda)^{-3}(1+\lambda)^{-1}T^{-3}\sigma_\varepsilon^2$$

where

$$A \equiv -2(1-\lambda-\ln\lambda)\lambda^{T+1}T - (1-\lambda)^2 T + 2(1-\lambda)(\ln\lambda)\lambda^{T+1}T^2 + 4\lambda(1-\lambda^T).$$

Observe that, $A < 4\lambda(1-\lambda^T) - (1-\lambda)^2 T < 4\lambda - (1-\lambda)^2 T$, so that $A < 0$ for all $T \geq T(\lambda) \equiv 4\lambda(1-\lambda)^{-2}$.

C Proof of Proposition 3

First let $\lambda^* = \frac{\sqrt{37}-1}{12} \approx .424$. λ^* is the (unique) value of λ at which $MSE_\infty^{est}(2) = MSE_\infty^{est}(3)$. Then $T(\lambda^*) \approx 5.1$. By Proposition 2, we know that $\text{MSE}_\infty^{est}(T)$ decreases with T for all $T \geq 6$. The question is what happens for $T \leq 5$. First observe that the inequality $MSE_\infty^{est}(T+1) - MSE_\infty^{est}(T) > 0$ is a polynomial in λ, for given T. One can then verify that for all $\lambda \leq \lambda^*$,[7]

$$MSE_\infty^{est}(6) < MSE_\infty^{est}(5) < MSE_\infty^{est}(4) < MSE_\infty^{est}(3) \leq MSE_\infty^{est}(2) < MSE_\infty^{est}(1).$$

[7] I used the "Inequality Solve" package in *Mathematica* Version 3.0 to solve algebraically for these and all of the succeeding polynomial inequalities.

This proves that the $\mathrm{MSE}^{est}_{\infty}(T)$ decreases with T for all $\lambda \leq \lambda^*$.

Note that $MSE^{est}_{\infty}(1) < MSE^{est}_{\infty}(2)$ iff $\lambda > .5$ (this may be easily verified directly by using (3)). Now consider the case when $\lambda \in (.424, .428]$. Proposition 2 tells us that the MSE decreases for all $T > 5$ in this interval of λ. It is also possible to verify that for all $\lambda \in (.424, .428]$, $MSE^{est}_{\infty}(6) < MSE^{est}_{\infty}(5) < MSE^{est}_{\infty}(4) \leq MSE^{est}_{\infty}(3)$ and $MSE^{est}_{\infty}(2) < MSE^{est}_{\infty}(3)$. Since we already know that $MSE^{est}_{\infty}(2) < MSE^{est}_{\infty}(1)$, the optimal T can be computed by comparing $MSE^{est}_{\infty}(2)$ and $MSE^{est}_{\infty}(T \to \infty)$ and it turns out that $MSE^{est}_{\infty}(T \to \infty) < MSE^{est}_{\infty}(2)$ iff $\lambda < .5$ (again this may be verified directly by using (3)). This proves that the optimal $T \to \infty$ when $\lambda \in (.424, .428]$.

Repeating the same type of arguments for neighbouring intervals like $(.428, .446], (.446, .466]$, $(.466, .486]$, and $(.486, .5]$, it can be shown that the optimal $T \to \infty$. Note that all these intervals arise out of comparing the MSE associated with adjacent memory lengths.

D Proof of Proposition 4

When $\lambda > .5$, it has been shown that $MSE^{est}_{\infty}(1) < MSE^{est}_{\infty}(2)$. Consider now the interval $(.5, .504]$ of λ. In this case one can show that the MSE increases monotonically from $T = 1$ to $T = 7$ and $MSE^{est}_{\infty}(8) \leq MSE^{est}_{\infty}(7)$. On the other hand, using Proposition 2, we know that the MSE decreases with T thereafter. Consequently, the optimal T can be computed by comparing $MSE^{est}_{\infty}(1)$ with $MSE^{est}_{\infty}(T \to \infty)$. But we already know that it can't be optimal to use $T \to \infty$ so that $T = 1$ is optimal when $\lambda \in (.5, .504]$. When $\lambda \in (.504, .521]$, it can similarly be shown that the MSE increases from $T = 1$ to $T = 8$ and, thereafter, decreases with T so that the optimal T is 1. One can continue in this fashion and look at higher intervals of λ. Thus, when $\lambda = .88$, $T(\lambda) = 250$. Proposition 2 tells us that the MSE decreases for all $T \geq 250$ and it is possible to show that the MSE increases from $T = 1$ to $T = 250$.

References

[1] Abraham, B., and J. Ledolter, 1983, Statistical Methods for Forecasting (Wiley-Interscience, New York).

[2] Evans, G.W. and S. Honkapohja, 1993. Adaptive Forecasts, Hysteris, and Endogenous Fluctuations. Economic Review Federal Reserve Bank of San Francisco 1, 3-13.

[3] Evans, G.W., and S. Honkapohja, 2001, Learning and Expectations in Macroeconomics (Princeton University Press, Princeton, New Jersey).

[4] Evans, G.W. and G. Ramey, 1998. Adaptive Expectations, Underparametrization and the Lucas Crique. mimeo.

[5] Hamilton, J., 1994, Time Series Analysis (Princeton University Press, Princeton, New Jersey).

[6] Hommes, C. H. and G. Sorger, 1998. Consistent Expectations Equilibria. Macroeconomic Dynamics 2, 287-321.

[7] Honkapohja, S. and K. Mitra, 2003. Learning with Bounded Memory in Stochastic Models. Journal of Economic Dynamics & Control 27, 1437-1457.

[8] Lucas, R. E., 1973. Some International Evidence on Output-Inflation Tradeoffs. American Economic Review, 63, 326-334.

[9] Lucas, R. E., 1976. Econometric Policy Evaluation: A Critique. Journal of Monetary Economics 1 (2), 19-46.

[10] Muth, J., 1961. Rational Forecasts and the theory of price movements. Econometrica 29, 315-335.

[11] Sargent, T. J., 1999, The Conquest of American Inflation (Princeton University Press, Princeton, New Jersey).